

CLAIMS

I claim:

1. A method for describing the goal of a data mining operation, the method comprising

5 providing a user interface having a control for receiving natural language input;

receiving natural language input describing the goal of the data mining operation from the control on the user interface.

10087240, 030102 10 2. The method for describing the goal of a data mining operation having a dependent variable according to claim 1, further comprising:

sending the natural language input to a text parser.

15 3. The method for describing the goal of a data mining operation having a dependent variable according to claim 2, wherein the text parser is available to identify keywords, the text parser is available to use Bayesian networks for lexical analysis to calculate maximum a posteriori probabilities for candidate target fields.

20 4. The method for describing the goal of a data mining operation having a dependent variable according to claim 2 further comprising:

identifying keywords with the text parser;

25 using Bayesian networks for lexical analysis of the natural language input with identified keywords;

providing a database having fields containing data;

selecting a field from the database as the target field,

using the results of the lexical analysis to calculate a maximum a posteriori probability that the target field is the dependent variable.

5 5. The method for describing the goal of a data mining operation according to claim 4 wherein the database fields have names, the method further comprising comparing the target field name with the result of the lexical analysis.

10 6. The method for describing the goal of a data mining operation according to claim 4 wherein the database fields have descriptions, the method further comprising comparing the target field description with the result of the lexical analysis.

15 7. The method for describing the goal of a data mining operation according to claim 4 further comprising:
 identifying candidate fields, the candidate fields being relatively more likely to be the dependent variable than other fields in the database;
 displaying the candidate fields;
 receiving selection input defining the dependent variable
20 based on the candidate fields.

8. The method for describing the goal of a data mining operation according to claim 7, wherein the selection input identifies one candidate field as the dependent variable.

25 9. The method for describing the goal of a data mining operation according to claim 7, wherein the selection input specifies a formula combining candidate fields to define the true independent variable.

10. The method according to claim 2, wherein the user interface resides on a client system and the text parser resides on a server system.

11. The method according to claim 2 wherein the user interface and the text parser reside on the same system.

12. A method in a computer system for communicating results of a data mining operation, the method comprising:

identifying key performance results;

10 providing a user interface having a control for communicating information;

communicating a natural language description of the key performance results using the control on the user interface.

13. The method in a computer system for communicating results of a data mining operation according to claim 12 further comprising

15 providing a robust data model comprising each algorithm used, each algorithm's parameters, each algorithm's performance results, and input/output specification with time tag; and

20 providing as part of the user interface text templates for communicating the key performance results.

14. The method in a computer system for communicating results of a data mining operation according to claim 13 further comprising

25 providing as part of the user interface a plurality text templates for communicating the key performance results;

selecting one text template from among the plurality of text templates for communicating the key performance results, whereby the user interface does not display the same text template for every data mining operation.

5 15. The method according to claim 14 wherein the user interface is provided on a client system and the data model is provide on a server.

16. The method according to claim 14 wherein the user interface and the client system are both contained on a
10 general-purpose computer.

17. A method in a computer system for controlling a data mining operation, the method comprising:

receiving problem specification input determining a data mining operation goal, wherein the input data determining a
15 data mining operation goal is the only input required by the data mining application.

18. The method according to claim 17 wherein the problem specification input is a formal definition based on a data model.

20 19. The method according to claim 17 wherein the problem specification input is natural language data.

20. The method according to claim 17 or claim 18 further comprising:

identifying key performance results;

25 providing a user interface having a control for communicating information;

communicating a natural language description of the key performance results using the control on the user interface.

21. A data mining application user interface comprising:

5 a control that receives natural language input describing the goal of a data mining operation; and

an interface that sends the natural language input to a text parser.

22. The data mining application user interface according
10 to claim 21, wherein the text parser is available to look for keywords, the text parser is available to perform lexical analysis using Bayesian networks, and the text parser is available to calculate maximum a posteriori probabilities for candidate target fields by comparing the results of the
15 lexical analysis with the table-space field names.

23. The data mining application user interface according to claim 22, wherein the input data determining a data mining operation goal is the only input required by the data mining application.

20 24. A computer data signal stream for communicating the goal of a data mining operation, the data signal stream comprising:

25 natural language input data describing the goal of the data mining operation, the natural language input data being available for lexical analysis to identify at least one candidate data field;

problem specification data which specifies a goal of the data mining operation based on the at least one candidate data field identified by lexical analysis.

25. A computer data signal stream for controlling a data mining operation, the data signal stream consisting essentially of input data specifying the goal of the data mining operation, whereby no additional input is required to
5 obtain useful results.

26. An article of manufacture for a data mining application, the data mining application being available to perform a data mining operation on a database having fields, the data mining operation based on a dependent variable, the
10 article of manufacture comprising a computer readable medium, the computer readable medium containing:

computer program code that provides for receiving natural language data describing the goal of a data mining operation;

15 computer program code that provides for sending the natural language data to a text parser;

computer program code that provides for performing a lexical analysis of the natural language data using a Bayesian network;

20 computer program code that compares results of the lexical analysis to a database field to calculate a maximum a posteriori probability that the database field is the dependent variable;

25 computer program code that outputs the identity of candidate database fields more likely than other database fields to be the dependent variable; and

computer program code that provides for receiving problem specification data based on the candidate database fields.

27. An article of manufacture for a data mining application, the article of manufacture comprising a computer readable medium containing

a plurality of natural language text templates for communicating the key performance results;

computer program code that selects one text templates from among the plurality of text templates for communicating the key performance results, whereby the user interface does not display the same text template for every data mining operation.

28. An article of manufacture for a data mining application, the article of manufacture comprising a computer readable medium containing computer program code that provides for receiving input determining a data mining operation goal, wherein the input determining a data mining operation goal is the only input required by the data mining application.

29. An article of manufacture for a data mining application, the article of manufacture comprising a computer readable medium containing computer program code selected from the group consisting of: computer program code that receives natural language text providing a data mining operation goal; computer program code that displays key data mining performance results in natural language text; and computer program code that receives input providing a data mining operation goal, wherein the input providing a data mining operation goal is the only input required by the data mining application.

30. A user control method for a data mining application, the user control method comprising:

specifying a goal of data mining in natural language text; and

displaying key data mining performance results in
natural language text.

31. The method for providing user control of a data
mining application according to claim 29, wherein the
5 specifying a data mining goal is the only user action
required, and further comprising an interrupt mechanism to
display intermediate results.

32. The control method for providing user control of a
data mining application according to claim 29 wherein
10 specifying a goal of data mining in natural language text
further comprises:

receiving natural language text describing a data mining
problem, wherein a data mining problem includes at
least one dependent variable,
15 performing lexical analysis on the natural language text
with a Bayesian network, and
recommending a small number of fields relatively likely
to be candidates for the at least one dependent
variable of the data mining operation goal.

20 33. The method for providing user control of a data
mining application according to claim 29 wherein specifying a
goal of data mining in natural language text further
comprises:

providing a set of named fields,
25 receiving natural language text describing a data mining
problem, wherein a data mining problem includes at
least one dependent variable,
identifying key words in the natural language text,
performing lexical analysis on the natural language text
30 with a Bayesian network,

calculating maximum a posteriori probabilities for
fields by comparing lexical analysis results with
field names, and
recommending a small number of fields relatively likely
to be candidates for the at least one dependent
variable of the data mining operation goal,
communicating the fields relatively likely to be
candidates to the user,
receiving additional input identifying the dependent
variable,
for each target candidate, ranking input features based
on their level of contribution to the expected data
mining performance.

34. The method for providing user control of a data
mining application according to claim 29 wherein displaying
key data mining performance results in natural language text

35. A problem specification method for mapping a data
mining goal expressed in natural language to data fields, the
method comprising:

providing a set of fields having field names,
receiving natural language text describing a data mining
operation goal, wherein a data mining operation goal
includes at least one dependent variable,
identifying key words in the natural language text,
performing lexical analysis on the natural language text
with a Bayesian network,
calculating maximum a posteriori probabilities for
fields by comparing lexical analysis results with
field names,
recommending a small number of fields relatively likely
to be candidates for the at least one dependent
variable of the data mining operation goal,

communicating the fields relatively likely to be
candidates to the user,
receiving additional user input specifying the dependent
variable, and

- 5 for each target candidate, ranking input features based
on their level of contribution to the expected data
mining performance.